

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-282827

(43)Date of publication of application : 12.10.2001

(51)Int.Cl.

G06F 17/30

(21)Application number : 2000-093657

(71)Applicant : JUST SYST CORP

(22)Date of filing : 30.03.2000

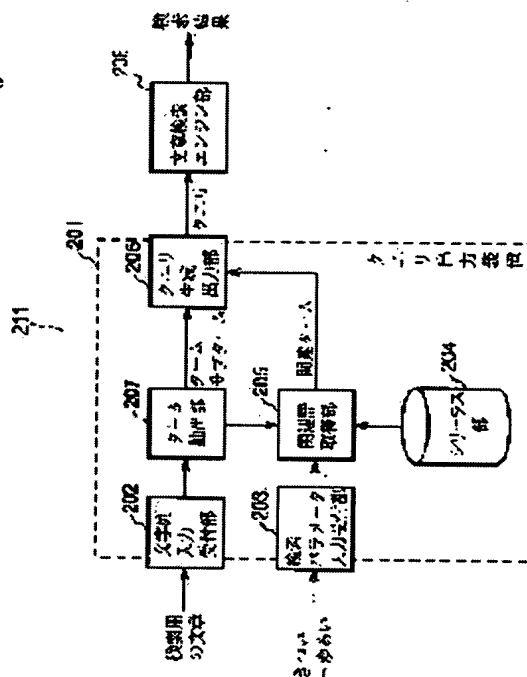
(72)Inventor : AJIKI TOSHIHIRO

(54) DEVICE OF RETRIEVAL FOR DOCUMENT, DEVICE OF OUTPUT FOR QUERY, METHOD OF RETRIEVAL FOR DOCUMENT, METHOD OF OUTPUT FOR QUERY AND STORAGE MEDIUM

(57)Abstract:

PROBLEM TO BE SOLVED: To provide a device of retrieval for a document and so on which retrieves the document, making an appropriate query by using a parameter.

SOLUTION: A locus of acceptance 202 for input of character accepts input of characters. A locus of extraction 207 for draw of terms extracts terms gotten by analyzing character roughly. A locus of acceptance 203 for input of retrieval parameter accepts input of retrieval parameter. A locus of thesaurus 204 memorizes more than one term corresponding to the retrieval parameter to each of plural terms. A locus of acquisition 205 for linking terms obtains memorized linking terms corresponding to the retrieval parameter accepted to each of terms extracted from the locus of thesaurus 204. A locus of output 206 for generating a query generates and put up a query containing extracted terms and linking terms obtained to each of extracted terms. A locus of engine 208 for retrieval document retrieves a document by appropriate query.



LEGAL STATUS

[Date of request for examination] 30.03.2000

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 3444592

[Date of registration] 27.06.2003

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開2001-282827

(P2001-282827A)

(43) 公開日 平成13年10月12日 (2001. 10. 12)

(51) Int.Cl.⁷

G 0 6 F 17/30

識別記号

3 2 0

1 7 0

F I

G 0 6 F 17/30

テーマコード* (参考)

3 2 0 D 5 B 0 7 5

1 7 0 A

審査請求 有 請求項の数15 O L (全 10 頁)

(21) 出願番号 特願2000-93657(P2000-93657)

(22) 出願日 平成12年3月30日 (2000. 3. 30)

(71) 出願人 390024350

株式会社ジャストシステム

徳島県徳島市沖浜東3-46

(72) 発明者 安食 敏宏

徳島県徳島市川内町平石若松108-4 株

式会社ジャストシステム内

(74) 代理人 100095407

弁理士 木村 満 (外1名)

Fターム(参考) 5B075 ND03 NK02 NK35 PP02 PP03

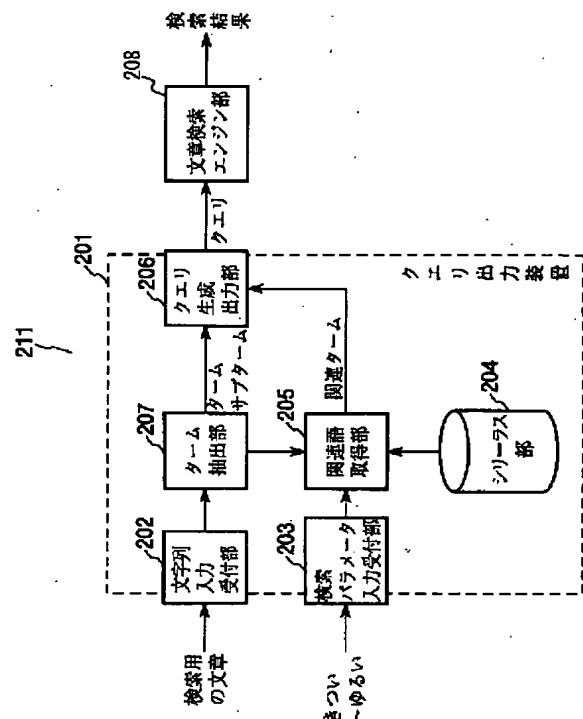
PP12 PP26 PQ02 QP03

(54) 【発明の名称】 文書検索装置、クエリ出力装置、文書検索方法、クエリ出力方法、並びに、記録媒体

(57) 【要約】

【課題】 パラメータを用いて適切なクエリを生成して文書を検索する文書検索装置等を提供する。

【解決手段】 文字列入力受付部202は、文字列の入力を受け付け、ターム抽出部207は、受け付けられた文字列を形態素解析し、これに含まれるタームを抽出し、検索パラメータ入力受付部203は、検索パラメータの入力を受け付け、シソーラス部204は、複数のタームのそれぞれに対して、検索パラメータに対応付けて関連タームを1つ以上記憶し、関連語取得部205は、シソーラス部204から、抽出されたタームのそれぞれに対して受け付けられた検索パラメータに対応付けられて記憶された関連タームを取得し、クエリ生成出力部206は、抽出されたタームと、そのそれぞれに対して取得された関連タームと、を含むクエリを生成して出力し、文書検索エンジン部208は、当該クエリを用いて文書を検索する。



【特許請求の範囲】

【請求項1】文字列の入力を受け付ける文字列入力受付部と、

検索パラメータの入力を受け付ける検索パラメータ入力受付部と、

複数の文字列のそれぞれに対して、検索パラメータに対応付けて関連文字列を1つ以上記憶するシソーラス部と、

前記シソーラス部から、前記受け付けられた文字列に対して前記受け付けられた検索パラメータに対応付けられて記憶された関連文字列を取得する関連語取得部と、

前記受け付けられた文字列と前記取得された関連文字列により文書を検索する文書検索部と、を備えることを特徴とする文書検索装置。

【請求項2】タームを含むクエリを文書に対して適用して得られた得点が所定の範囲に含まれる文書を獲得する文書検索エンジンに与えるクエリを出力するクエリ出力装置であって、

文字列の入力を受け付ける文字列入力受付部と、

前記受け付けられた文字列を形態素解析し、これに含まれるタームを抽出するターム抽出部と、

検索パラメータの入力を受け付ける検索パラメータ入力受付部と、

複数のタームのそれぞれに対して、検索パラメータに対応付けて関連タームを1つ以上記憶するシソーラス部と、

前記シソーラス部から、前記抽出されたタームのそれぞれに対して前記受け付けられた検索パラメータに対応付けられて記憶された関連タームを取得する関連語取得部と、

前記抽出されたタームと、そのそれぞれに対して前記取得された関連タームと、を含むクエリを生成して出力するクエリ生成出力部と、を備えることを特徴とするクエリ出力装置。

【請求項3】前記生成されるクエリは、前記受け付けられた検索パラメータに対応付けられてあらかじめ定められた重みにより、前記抽出されたタームと前記取得された関連タームとを重み付けすることを特徴とする請求項2に記載のクエリ出力装置。

【請求項4】前記生成されるクエリにおいて、前記抽出されたタームに対する重み付けの重みと、前記取得された関連タームに対する重み付けの重みと、は異なることを特徴とする請求項3に記載のクエリ出力装置。

【請求項5】クエリを出力する請求項2から4のいずれか1項に記載のクエリ出力装置と、

前記出力されたクエリを文書に対して適用して得られた得点が所定の範囲に含まれる文書を獲得して出力する文書検索エンジン部と、を備えることを特徴とする文書検索装置。

【請求項6】文字列の入力を受け付ける文字列入力受付

工程と、

検索パラメータの入力を受け付ける検索パラメータ入力受付工程と、

複数の文字列のそれぞれに対して検索パラメータに対応付けて記憶された1つ以上の関連文字列から、前記受け付けられた文字列に対して前記受け付けられた検索パラメータに対応付けられて記憶された関連文字列を取得する関連語取得工程と、

前記受け付けられた文字列と前記取得された関連文字列により文書を検索する文書検索工程と、を備えることを特徴とする文書検索方法。

【請求項7】タームを含むクエリを文書に対して適用して得られた得点が所定の範囲に含まれる文書を獲得する文書検索エンジンに与えるクエリを出力するクエリ出力方法であって、

文字列の入力を受け付ける文字列入力受付工程と、

前記受け付けられた文字列を形態素解析し、これに含まれるタームを抽出するターム抽出工程と、

検索パラメータの入力を受け付ける検索パラメータ入力受付工程と、

複数のタームのそれぞれに対して検索パラメータに対応付けて記憶された1つ以上の関連タームから、前記抽出されたタームのそれぞれに対して前記受け付けられた検索パラメータに対応付けられて記憶された関連タームを取得する関連語取得工程と、

前記抽出されたタームと、そのそれぞれに対して前記取得された関連タームと、を含むクエリを生成して出力するクエリ生成出力工程と、を備えることを特徴とするクエリ出力方法。

【請求項8】前記生成されるクエリは、前記受け付けられた検索パラメータに対応付けられてあらかじめ定められた重みにより、前記抽出されたタームと前記取得された関連タームとを重み付けすることを特徴とする請求項7に記載のクエリ出力方法。

【請求項9】前記生成されるクエリにおいて、前記抽出されたタームに対する重み付けの重みと、前記取得された関連タームに対する重み付けの重みと、は異なることを特徴とする請求項8に記載のクエリ出力方法。

【請求項10】請求項7から9のいずれか1項に記載のクエリ出力方法を使用してクエリを出力するクエリ出力工程と、

前記出力されたクエリを文書に対して適用して得られた得点が所定の範囲に含まれる文書を獲得して出力する文書検索エンジン工程と、を備えることを特徴とする文書検索方法。

【請求項11】コンピュータを、

文字列の入力を受け付ける文字列入力受付部、

検索パラメータの入力を受け付ける検索パラメータ入力受付部、

複数の文字列のそれぞれに対して、検索パラメータに対

応付けて関連文字列を1つ以上記憶するシソーラス部、前記シソーラス部から、前記受け付けられた文字列に対して前記受け付けられた検索パラメータに対応付けられて記憶された関連文字列を取得する関連語取得部、および、

前記受け付けられた文字列と前記取得された関連文字列とにより文書を検索する文書検索部として機能させることを特徴とするプログラムを記録したコンピュータ読取可能な記録媒体。

【請求項12】コンピュータを、タームを含むクエリを文書に対して適用して得られた得点が所定の範囲に含まれる文書を獲得する文書検索エンジンに与えるクエリを出力するクエリ出力装置として機能させるプログラムであって、

前記プログラムは、前記コンピュータを、文字列の入力を受け付ける文字列入力受付部、前記受け付けられた文字列を形態素解析し、これに含まれるタームを抽出するターム抽出部、検索パラメータの入力を受け付ける検索パラメータ入力受付部、

複数のタームのそれぞれに対して、検索パラメータに対応付けて関連タームを1つ以上記憶するシソーラス部、前記シソーラス部から、前記抽出されたタームのそれぞれに対して前記受け付けられた検索パラメータに対応付けられて記憶された関連タームを取得する関連語取得部、および、

前記抽出されたタームと、そのそれぞれに対して前記取得された関連タームと、を含むクエリを生成して出力するクエリ生成出力部として機能させることを特徴とするプログラムを記録したコンピュータ読取可能な記録媒体。

【請求項13】前記プログラムは、前記コンピュータにおいて、

前記生成されるクエリは、前記受け付けられた検索パラメータに対応付けられてあらかじめ定められた重みにより、前記抽出されたタームと前記取得された関連タームとを重み付けするように機能させることを特徴とする請求項12に記載の記録媒体。

【請求項14】前記プログラムは、前記コンピュータにおいて、

前記生成されるクエリにおいて、前記抽出されたタームに対する重み付けの重みと、前記取得された関連タームに対する重み付けの重みと、は異なるように機能させることを特徴とする請求項13に記載の記録媒体。

【請求項15】コンピュータを、クエリを出力する請求項2から4のいずれか1項に記載のクエリ出力装置、および、

前記出力されたクエリを文書に対して適用して得られた得点が所定の範囲に含まれる文書を獲得して出力する文書検索エンジン部として機能させることを特徴とするプ

ログラムを記録したコンピュータ読取可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、検索にパラメータを与え、キーワードからパラメータに応じた関連語を得て、これらによりクエリを生成して文書を検索するのに好適な文書検索装置、クエリ出力装置、文書検索方法、クエリ出力方法、並びに、これらを実現するためのプログラムを記録したコンピュータ読取可能な記録媒体に関する。

【0002】

【従来の技術】WWW（World Wide Web）技術の発展にともない、多数の文書から所望の文書を検索する文書検索技術はますますその重要性を増している。

【0003】このような文書検索では、従来、以下のような手法が用いられていた。すなわち、ユーザは、所望の文書を検索するためのキーワードを指定するか、あるいは、文書を検索するための条件をより詳細に指定したクエリを入力する。

【0004】すると、文書検索エンジンは、文書中に当該キーワードが含まれる個数や、ベクトル空間法によりクエリにより指定された条件と文書との適合度を求め、これらを元に当該文書の得点を計算する。

【0005】最後に、ユーザに対して、得点の高い文書を検索結果として順に提示する。

【0006】なお、ベクトル空間法では、クエリに含まれる検索用のキーワード（以下「ターム」という。）に対して適宜重みを付ける手法や、文書検索エンジンの検索範囲に含まれる文書に対して、それぞれ、当該文書から抽出されたターム（以下「検索用ターム」という。）を別途記憶しておき、得点計算を高速化する手法も提案されている。

【0007】

【発明が解決しようとする課題】しかしながら、指定されたキーワードが用意された検索用タームでない場合には、検索結果として有意な結果が得られないおそれがある。

【0008】一方で、このような検索用タームを一般ユーザがあらかじめ自分で調査するのは繁雑であり、また、適切な検索用タームを選択できないおそれもある。

【0009】このような検索用タームは、一般ユーザが指定したキーワードから自動的に選択されることが強く望まれている。

【0010】本発明が解決しようとする課題は、検索にパラメータを与え、キーワードからパラメータに応じた関連語を得て、これらによりクエリを生成して文書を検索することにある。

【0011】

【課題を解決するための手段】上記課題を解決するため、本発明の第1の観点に係る文書検索装置は、文字列

入力受付部と、検索パラメータ入力受付部と、シソーラス部と、関連語取得部と、文書検索部と、を備えるように構成する。

【0012】ここで、文字列入力受付部は、文字列の入力を受け付ける。

【0013】一方、検索パラメータ入力受付部は、検索パラメータの入力を受け付ける。

【0014】さらに、シソーラス部は、複数の文字列のそれぞれに対して、検索パラメータに対応付けて関連文字列を1つ以上記憶する。

【0015】そして、関連語取得部は、前記シソーラス部から、前記受け付けられた文字列に対して前記受け付けられた検索パラメータに対応付けられて記憶された関連文字列を取得する。

【0016】一方、文書検索部は、前記受け付けられた文字列と前記取得された関連文字列とにより文書を検索する。

【0017】本発明により、あらかじめ定めた検索用タームを、一般ユーザが指定した簡単なパラメータによって自動的に取得して、これらにより文書の検索を適切に行うことができる。

【0018】本発明の第2の観点に係るクエリ出力装置は、タームを含むクエリを文書に対して適用して得られた得点が所定の範囲に含まれる文書を獲得する文書検索エンジンに与えるクエリを出力し、文字列入力受付部と、ターム抽出部と、検索パラメータ入力受付部と、シソーラス部と、関連語取得部と、クエリ生成出力部と、を備えるように構成する。

【0019】ここで、文字列入力受付部は、文字列の入力を受け付ける。

【0020】一方、ターム抽出部は、前記受け付けられた文字列を形態素解析し、これに含まれるタームを抽出する。

【0021】さらに、検索パラメータ入力受付部は、検索パラメータの入力を受け付ける。

【0022】そして、シソーラス部は、複数のタームのそれぞれに対して、検索パラメータに対応付けて関連タームを1つ以上記憶する。

【0023】一方、関連語取得部は、前記シソーラス部から、前記抽出されたタームのそれぞれに対して前記受け付けられた検索パラメータに対応付けられて記憶された関連タームを取得する。

【0024】さらに、クエリ生成出力部は、前記抽出されたタームと、そのそれぞれに対して前記取得された関連タームと、を含むクエリを生成して出力する。

【0025】本発明により、一般ユーザが入力した普通の文章から、あらかじめ定めた検索用タームを、一般ユーザが指定した簡単なパラメータによって、自動的に取得して、文書の検索を行うためのクエリを適切に生成して出力することができる。

【0026】また、本発明のクエリ出力装置において、前記生成されるクエリは、前記受け付けられた検索パラメータに対応付けられてあらかじめ定められた重みにより、前記抽出されたタームと前記取得された関連タームとを重み付けするように構成することができる。

【0027】本発明により、あらかじめ定めた検索用タームのそれぞれに、一般ユーザが指定したパラメータに基づいて重み付けを行い、一般ユーザの所望の範囲の文書を検索するためのクエリを適切に生成して出力することができる。

【0028】また、本発明のクエリ出力装置において、前記生成されるクエリにおいて、前記抽出されたタームに対する重み付けの重みと、前記取得された関連タームに対する重み付けの重みと、は異なるように構成することができる。

【0029】本発明により、一般ユーザが入力した検索用の普通の文章に検索用タームが含まれる場合には、これの重みを大きくし、これに関連するタームの重みは小さくするようなクエリを適切に生成して出力することができる。

【0030】本発明の第3の観点に係る文書検索装置は、上記のクエリ出力装置と、文書検索エンジン部と、を備えるように構成する。

【0031】ここで、上記のクエリ出力装置はクエリを出力する。

【0032】一方、文書検索エンジン部は、前記出力されたクエリを文書に対して適用して得られた得点が所定の範囲に含まれる文書を獲得して出力する。

【0033】本発明により、一般ユーザが入力した検索用の普通の文章から、あらかじめ定めた検索用タームを、一般ユーザが指定した簡単なパラメータによって自動的に取得して、これらにより文書の検索を適切に行うことができる。

【0034】本発明の第4の観点に係る文書検索方法は、文字列入力受付工程と、検索パラメータ入力受付工程と、関連語取得工程と、文書検索工程と、を備えるように構成する。

【0035】ここで、文字列入力受付工程では、文字列の入力を受け付ける。

【0036】一方、検索パラメータ入力受付工程では、検索パラメータの入力を受け付ける。

【0037】さらに、関連語取得工程では、複数の文字列のそれぞれに対して検索パラメータに対応付けて記憶された1つ以上の関連文字列から、前記受け付けられた文字列に対して前記受け付けられた検索パラメータに対応付けられて記憶された関連文字列を取得する。

【0038】そして、文書検索工程では、前記受け付けられた文字列と前記取得された関連文字列とにより文書を検索する。

【0039】本発明の第5の観点に係るクエリ出力方法

10

20

30

40

50

は、タームを含むクエリを文書に対して適用して得られた得点が所定の範囲に含まれる文書を獲得する文書検索エンジンに与えるクエリを出力し、文字列入力受付工程と、ターム抽出工程と、検索パラメータ入力受付工程と、関連語取得工程と、クエリ生成出力工程と、を備えるように構成する。

【0040】ここで、文字列入力受付工程では、文字列の入力を受け付ける。

【0041】一方、ターム抽出工程では、前記受け付けられた文字列を形態素解析し、これに含まれるタームを抽出する。

【0042】さらに、検索パラメータ入力受付工程では、検索パラメータの入力を受け付ける。

【0043】そして、関連語取得工程では、複数のタームのそれぞれに対して検索パラメータに対応付けて記憶された1つ以上の関連タームから、前記抽出されたタームのそれぞれに対して前記受け付けられた検索パラメータに対応付けられて記憶された関連タームを取得する。

【0044】一方、クエリ生成出力工程では、前記抽出されたタームと、そのそれぞれに対して前記取得された関連タームと、を含むクエリを生成して出力する。

【0045】また、本発明のクエリ出力方法において、前記生成されるクエリは、前記受け付けられた検索パラメータに対応付けられてあらかじめ定められた重みにより、前記抽出されたタームと前記取得された関連タームとを重み付けするように構成することができる。

【0046】また、本発明のクエリ出力方法において、前記生成されるクエリにおいて、前記抽出されたタームに対する重み付けの重みと、前記取得された関連タームに対する重み付けの重みと、は異なるように構成することができる。

【0047】本発明の第6の観点に係る文書検索方法は、クエリ出力工程と、文書検索エンジン工程と、を備える。

【0048】ここで、クエリ出力工程では、上記クエリ出力方法を使用してクエリを出力する。

【0049】一方、文書検索エンジン工程では、前記出力されたクエリを文書に対して適用して得られた得点が所定の範囲に含まれる文書を獲得して出力する。

【0050】本発明の第7の観点に係るコンピュータ読取可能な記録媒体は、コンピュータを、文字列入力受付部、検索パラメータ入力受付部、シソーラス部、関連語取得部、および、文書検索部として機能させるプログラムを記録するように構成する。

【0051】ここで、文字列入力受付部は、文字列の入力を受け付ける。

【0052】一方、検索パラメータ入力受付部は、検索パラメータの入力を受け付ける。

【0053】さらに、シソーラス部は、複数の文字列のそれぞれに対して、検索パラメータに対応付けて関連文

字列を1つ以上記憶する。

【0054】そして、関連語取得部は、前記シソーラス部から、前記受け付けられた文字列に対して前記受け付けられた検索パラメータに対応付けられて記憶された関連文字列を取得する。

【0055】一方、文書検索部は、前記受け付けられた文字列と前記取得された関連文字列とにより文書を検索する。

【0056】本発明の第8の観点に係るコンピュータ読取可能な記録媒体は、コンピュータを、タームを含むクエリを文書に対して適用して得られた得点が所定の範囲に含まれる文書を獲得する文書検索エンジンに与えるクエリを出力するクエリ出力装置として機能させるプログラムを記録するように構成する。

【0057】ここで、前記プログラムは、前記コンピュータを、文字列入力受付部、ターム抽出部、検索パラメータ入力受付部、シソーラス部、関連語取得部、および、クエリ生成出力部として機能させる。

【0058】ここで、文字列入力受付部は、文字列の入力を受け付ける。

【0059】一方、ターム抽出部は、前記受け付けられた文字列を形態素解析し、これに含まれるタームを抽出する。

【0060】さらに、検索パラメータ入力受付部は、検索パラメータの入力を受け付ける。

【0061】そして、シソーラス部は、複数のタームのそれぞれに対して、検索パラメータに対応付けて関連タームを1つ以上記憶する。

【0062】一方、関連語取得部は、前記シソーラス部から、前記抽出されたタームのそれぞれに対して前記受け付けられた検索パラメータに対応付けられて記憶された関連タームを取得する。

【0063】さらに、クエリ生成出力部は、前記抽出されたタームと、そのそれぞれに対して前記取得された関連タームと、を含むクエリを生成して出力する。

【0064】また、本発明のコンピュータ読取可能な記録媒体に記録されるプログラムは、前記コンピュータにおいて、前記生成されるクエリは、前記受け付けられた検索パラメータに対応付けられてあらかじめ定められた重みにより、前記抽出されたタームと前記取得された関連タームとを重み付けするように機能させるように構成することができる。

【0065】また、本発明のコンピュータ読取可能な記録媒体に記録されるプログラムは、前記コンピュータにおいて、前記生成されるクエリにおいて、前記抽出されたタームに対する重み付けの重みと、前記取得された関連タームに対する重み付けの重みと、は異なるように機能させるように構成することができる。

【0066】本発明の第9の観点に係るコンピュータ読取可能な記録媒体に記録されるプログラムは、コンピュ

ータを、上記のクエリ出力装置、および、文書検索エンジン部として機能させるように構成する。

【0067】ここで、上記のクエリ出力装置はクエリを出力する。

【0068】一方、文書検索エンジン部は、前記出力されたクエリを文書に対して適用して得られた得点が所定の範囲に含まれる文書を獲得して出力する。

【0069】

【発明の実施の形態】以下、添付図面を参照して、本発明の実施の形態について説明する。

【0070】（概要構成）図1は、本発明のクエリ出力装置として機能する典型的な情報処理装置（汎用のコンピュータ、各種端末、携帯端末、携帯電話、ゲーム装置などの専用機器を含む）の概要構成を示す模式図である。以下、本図を参照して説明する。

【0071】情報処理装置101は、CPU（Central Processing Unit；中央処理ユニット）102によって制御される。情報処理装置101に電源を投入すると、CPU102は、ROM（Read Only Memory；読出専用メモリ）103に記憶されたIPL（Initial Program Loader；初期プログラムローダ）を実行する。

【0072】IPLは、ハードディスク104、FD（Floppy Disk；フロッピー（登録商標）ディスク）ドライブ110に装着されたFD、CD-ROM（Compact Disk ROM）ドライブ111に装着されたCD-ROMなどの記録媒体に記憶されたOS（Operating System；オペレーティング・システム）プログラムを読み出して実行するプログラムである。

【0073】OSを起動した後、CPU102は、キーボード105やマウス106などにより入力されたユーザの指示にしたがって、あるいは、ハードディスクなどにあらかじめ記述された設定ファイルの内容にしたがって、ハードディスクなどに記憶されたアプリケーションプログラムを実行する。

【0074】なお、携帯端末などの小規模な情報処理装置では、IPL自体がOSやアプリケーションプログラムとしての機能を果たすような実施形態も採用することができる。

【0075】CPU102は、プログラムの実行の際に、RAM（Random Access Memory）107を一時的な作業用記憶領域として用いる。このほか、一時的な作業用記憶領域として、CPU102内に設けられたレジスタやキャッシュ（図示せず）が使われる。

【0076】プログラムの実行に伴ない、ユーザに結果を報告したり、途中経過を見せるため、CPU102は、液晶ディスプレイやCRT（Cathode Ray Tube）などの表示装置108に当該情報を表示することができる。マウス106による指示操作では、マウス106を移動することにより、画面に表示されたカーソルが移動し、マウス106をクリックすることにより、カーソル

が指すメニュー項目を選択することができる。

【0077】情報処理装置101は、NIC（Network Interface Card）やモデムなどのインターフェース109を介してインターネットなどのコンピュータ通信網と通信を行うことができる。インターフェース109を介して他のサーバコンピュータ（文書検索エンジン）にクエリを送信したり、これから応答として送信された検索結果を受信することができる。また、インターフェース109を介して受信したプログラムを実行することもできる。

【0078】図2は、上記情報処理装置101が本発明のクエリ出力装置並びに文書検索装置として機能する場合の、各機能の概要構成を示す模式図である。以下、本図を参照して説明する。

【0079】クエリ出力装置201は、文字列入力受付部202と、検索パラメータ入力受付部203と、シソーラス部204と、関連語取得部205と、クエリ生成出力部206と、を備える。

【0080】さらに、本実施形態では、ユーザが普通の文章で検索用キーワードを入力できるようにするため、ターム抽出部207を用意する。なお、ユーザがキーワードを単語や語句単位で入力するような実施形態では、ターム抽出部207を用意しなくともよい。

【0081】さらに、本実施形態では、生成出力されたクエリにより、実際に文書の検索を行う文書検索エンジン部208を用意する。

【0082】文字列入力受付部202は、ユーザから検索用の自然な文章を表す文字列の入力を受け付ける。したがって、キーボード105やマウス106などの入力装置は、文字列入力受付部202として機能する。

【0083】ターム抽出部207は、与えられた文字列を形態素解析し、これに含まれるタームを抽出する。したがって、CPU102は、ターム抽出部207として機能する。

【0084】検索パラメータ入力受付部203は、たとえば、「大変ゆるい検索～ゆるい検索～普通の検索～きつい検索～大変きつい検索」のような5段階からいずれかを検索パラメータとしてユーザが選択した結果の入力を受け付ける。したがって、キーボード105やマウス106などの入力装置は、検索パラメータ入力受付部203として機能する。

【0085】シソーラス部204は、タームのそれぞれに対して、上記検索パラメータのそれぞれに対応付けて、関連する関連タームを1つ以上記憶する。したがって、ROM103、ハードディスク104、FDD110に装着されたFD、CD-ROMドライブ111に装着されたCD-ROM、インターフェース109を介して接続された他のコンピュータは、シソーラス部204として機能する。

【0086】関連語取得部205は、形態素解析の結果

抽出されて得られたタームのそれぞれについて、シソーラス部204から、現在の検索パラメータに基づいて、関連タームを取得する。したがって、CPU102は、関連語取得部205として機能する。

【0087】クエリ生成部206は、ターム抽出部207が抽出したタームと、関連語取得部20が取得した関連タームと、を含むクエリを生成して出力する。したがって、CPU102は、クエリ生成出力部206として機能する。

【0088】文書検索エンジン部208は、生成出力されたクエリにより、文書を検索する。したがって、CPU102が、これらの文書を記憶するROM103、ハードディスク104、FDD110に装着されたFD、CD-ROMドライブ111に装着されたCD-ROMと共働して、文書検索エンジン部208として機能する。

【0089】また、たとえばWWW検索サーバのように、インターフェース109を介して接続された他のコンピュータが、文書検索エンジン部208として機能する。この場合は、文書検索エンジン部208以外の部分

が、クエリ出力装置201として機能する。

【0090】このように、文書検索エンジン部208と、クエリ出力装置201とを、同じ情報処理装置101上に構成することも、異なる情報処理装置101上に構成することも、いずれも可能である。ただし、負荷分散や文書データベースの集中管理の観点からは、これらを分離した態様とすることが望ましい。

【0091】このほか、ハードディスク104、FDDドライブ110に装着されたFD、CD-ROMドライブ111に装着されたCD-ROM、ROM103、RAM107、および、インターフェース109を介して接続された他のコンピュータなど、これらコンピュータ読取可能な情報記憶装置に相当するものは、CPU102で実行されるプログラムを記録することにより、本発明の記録媒体として機能する。

【0092】上述のように、図1に示す情報処理装置101は、図2に示す本発明の文書検索装置211、並びに、クエリ出力装置201として機能する。これは、本発明のクエリ出力装置201、および、文書検索装置211の典型的な実施態様である。

【0093】一方、携帯端末等にて本発明の文書検索装置211、並びに、クエリ出力装置201を実現する場合は、外部記憶装置として、不揮発性半導体メモリ（フラッシュメモ리카ード）やバッテリーバックアップがされたRAMを使用することができる。この場合、必ずしもハードディスク104、FDDドライブ110、CD-ROMドライブ111等は必要ない。

【0094】（文書検索処理）図3は、図1に示す情報処理装置101を本発明の文書検索装置211として機能させる場合の文書検索処理の手順を示すフローチャー

トである。以下、本図を参照して説明する。

【0095】まず、CPU102は、ユーザから検索対象の文書を特定するための検索条件を示す文章の入力を受け付ける（ステップS301）。ユーザは、たとえば、「日本と中国の合併企業」のような自然な文章を検索条件として入力することができる。

【0096】また、この際に、合わせて検索パラメータの入力も受け付ける（ステップS302）。本実施形態では、検索のゆるさ〜きつさを5段階で入力することができる。

【0097】図4は、このような文章入力と検索パラメータの入力の両方を受け付ける入力フォームの表示例である。

【0098】入力フォーム401は、情報処理装置101の表示装置108に表示される。ユーザは、検索に用いる文章を、文章入力欄402に入力し、検索のゆるさ〜きつさをパラメータ用スライダ403を用いて入力する。これらを入力した後に、検索実行ボタン404を用いてクエリの生成、および、文書の検索の実行を指示する。

【0099】次に、CPU102は、入力された文字列を形態素解析して、これに含まれるタームを抽出する（ステップS303）。上記入力例の場合は、タームは形態素解析の結果は、「日本／と／中国／の／合併／企業」のように形態素解析される。ここで「／」は形態素の区切りを示す。ここからタームとして「日本」「中国」「合併／企業」を抽出する。

【0100】タームがほかのタームを含む場合には、これらも抽出する（ステップS304）。他のタームに含まれるタームを「サブターム」という。たとえば、「合併／企業」のサブタームは、「合併」と「企業」である。

【0101】さらに、これらタームやサブタームの同義語・類義語など、関連するターム（以下「関連ターム」という。）を、シソーラス部204を用いて取得する（ステップS305）。シソーラス部204には、検索パラメータのゆるさ〜きつさに応じて、各種の関連タームが記憶されている。

【0102】本実施形態では、あるタームと、この関連タームとの情報は、あらかじめハードディスク104などに記憶しておく。たとえば、「日本」の関連タームとして、きつい検索用には、「日本国」「ニッポン」「ジャパン」「Japan」などが記憶されている。ゆるい検索用には、「東アジア」「北海道」「九州」「四国」「本州」などが記憶されている。

【0103】さらに、CPU102は、タームと、形態素解析によって得られた当該タームの品詞情報や係り受けの情報と、検索パラメータと、から、クエリを生成して出力する（ステップS306）。

【0104】ここで生成されるクエリには、以下の情報

が含まれる。

- ・ターム。サブタームや関連タームを含む。
- ・当該タームの重み。ベクトル空間法で文書の得点を計算する際に用いる。
- ・必須ターム情報。当該タームが必須タームか否かを示す。

【0105】タームの重みや必須ターム情報の決定の手法については後述する。

【0106】なお、出力先としては、たとえば、RAM 107を採用することができる。

【0107】さらに、出力されたクエリを用いて、文書検索エンジン部208へ問い合わせ（ステップS307）、本処理を終了する。この詳細については、後述する。なお、本実施形態では、上述のように、同じ情報処理装置101上にクエリ出力装置201と文書検索エンジン部208とが実現されているが、異なる情報処理装置101上にこれらを配置してもよい。

【0108】（タームの例）たとえば、ユーザが「日本と中国の合併企業」を入力した場合に、抽出もしくは取得されるタームの様子を以下に示す。

【0109】「きつい検索」を選んだ場合には、以下の抽出済タームと関連タームからクエリを生成する。

- ・日本
- ・Japan、ジャパン、ニッポン（「日本」の関連ターム）
- ・中国
- ・China、チャイナ（「中国」の関連ターム）
- ・合併企業

【0110】「ゆるい検索」を選んだ場合には、以下の抽出済タームと関連タームからクエリを生成する。

- ・日本
- ・Japan、ジャパン、ニッポン（「日本」の関連ターム）
- ・北海道、本州、九州、四国（「日本」の関連ターム）
- ・中国（「中国」の関連ターム）
- ・China、チャイナ（「中国」の関連ターム）
- ・広州、香港、澳門（「中国」の関連ターム）
- ・合併企業
- ・合併（「合併企業」のサブターム）
- ・企業（「合併企業」のサブターム）

【0111】このように、検索パラメータの相違に基づいて、異なるタームを用いて検索用クエリを生成することにより、ユーザの所望の範囲の文書を検索結果として得ることが容易になる。

【0112】（問合せ処理）図5は、出力されたクエリを用いて、異なる情報処理装置101上に配置された文書検索エンジン部208へ問い合わせを行う問合せ処理の流れを示すフローチャートである。

【0113】CPU102は、ステップS307におい

て、RAM107に出力されたクエリを、インターフェース109を介して文書検索エンジン部208へ送信する（ステップS601）。

【0114】文書検索エンジン部208は、当該クエリを受信し（ステップS602）、当該クエリに基づいて、各文書の得点を計算し（ステップS603）、上位の所定の数の文書を取得し（ステップS604）、これらを検索結果として送信する（ステップS605）。

【0115】CPU102は、インターフェース109を介して、検索結果の文書を受信し（ステップS606）、当該文書を表示装置108に表示し、あるいは、RAM107やハードディスク104などの記憶装置に出力して（ステップS607）、本処理を終了する。

【0116】このように、ユーザは、自然な文章を入力することで検索の条件を指定することができる一方、検索パラメータを入力することで、所望の範囲に含まれる検索結果を得ることができる。

【0117】（重みの決定）クエリ生成部206、および、ステップS306においては、タームと、形態素解析によって得られた当該タームの品詞情報や係り受けの情報と、検索パラメータと、から、クエリを生成する。以下では、このクエリにおけるタームの重みの決定手法の一例について説明する。

【0118】日本語では名詞同士の係り受けや複合名詞が多用されるため、文書検索においてもこれらを考慮する必要がある。そこで、これらを検索結果の傾向から以下のように類型化して、各タームの重みを決定する。

【0119】第1の類型は、一般名詞と一般名詞が接続（助詞「の」を挟んで連なる場合を含む。以下同様。）する場合である。この場合は、前に配置された一般名詞が主題となる場合が多い。たとえば、「食事／療法」「食事／の／療法」「ゴルフ／トーナメント」「ゴルフ／の／トーナメント」の場合、主題となるのはそれぞれ「食事」「ゴルフ」である。そこで、これら主題となる前に配置されたタームの重みを大きくする。

【0120】第2の類型は、サブタームを含むタームの場合である。サブタームとタームとで、意味が大きく異なる場合は、サブタームの重みを低くし、全体のタームの重みを高くする。たとえば、「地方／銀行」「就職／協定」では、これら全体の重みを高くし、「地方」「銀行」「就職」「協定」の重みは低くする。

【0121】本実施形態では、これらの重み決定の手法を採用するが、自然言語による研究の成果に応じ、他の重みの決定のためのヒューリスティクスを採用することができる。また、他国語についても同様の処理を行うことができる。

【0122】（必須ターム情報の決定）クエリ生成部206、および、ステップS306においては、タームと、形態素解析によって得られた当該タームの品詞情報や係り受けの情報と、検索パラメータと、から、クエリ

を生成する。以下では、このクエリにおける必須ターム情報の決定手法の一例について説明する。

【0123】必須タームか否かは、そのタームが主題であるか否かに関わる。そこで、上記の重み決定の類型において、「重みを高くする」としたタームを必須タームとすることができる。

【0124】また、上記第2の類型において、サブターム「地方」「銀行」「就職」「協定」を用いずにクエリを作成してもよい。

【0125】（検索パラメータによる調整）本実施形態では、ユーザが入力した検索パラメータに応じてタームの重みや必須ターム情報を決定する。図4に示す入力フォーム401では、スライダ403により「きつい検索」から「ゆるい検索」まで検索パラメータを数段階の値として入力することができる。

【0126】「きつい検索」の場合には、重みの高低の差を増し、あるいは、必須タームの割合を増す。一方、「ゆるい検索」の場合には、重みの高低の差を減らし、あるいは、必須タームの割合を減らす。

【0127】たとえば、「格安／旅行」（後に配置されたタームが主題となる場合）について、「ゆるい検索」から「きつい検索」までクエリに含まれる情報を順に列挙すると以下ようになる。

- ・「格安」は重み小、「旅行」は重み小
- ・「格安」は重み小、「旅行」は重み大
- ・「格安」は重み小、「旅行」は重み大かつ必須
- ・「格安」は重み大、「旅行」は重み大かつ必須
- ・「格安」は重み大かつ必須、「旅行」は重み大かつ必須

【0128】さらに、本実施形態では、「きつい検索」の場合には、抽出された各タームに対してきつい検索用の関連タームとして記憶されたものを採用し、「ゆるい検索」の場合には、ゆるい検索用の関連タームとして記憶されたものを採用する。これらにも、上記と同様に重み付けをすることができる。また、あらかじめ重みをシソーラス部204に記憶させておいてもよい。

【0129】これらの検索パラメータと生成されるクエリとの関係についても、適宜変更が可能であり、変更した実施形態も本発明の範囲に含まれる。

【0130】

【発明の効果】本発明により、検索にパラメータを与え、キーワードからパラメータに応じた関連語を得て、

これらによりクエリを生成して文書を検索するのに好適な文書検索装置、クエリ出力装置、文書検索方法、クエリ出力方法、並びに、これらを実現するためのプログラムを記録したコンピュータ読取可能な記録媒体を提供することができる。

【図面の簡単な説明】

【図1】本発明の文書検索装置、並びに、クエリ出力装置として機能する典型的な情報処理装置の概要構成を示す模式図である。

【図2】本発明の文書検索装置、並びに、クエリ出力装置の各機能の概要構成を示す模式図である。

【図3】本発明の文書検索処理の手順を示すフローチャートである。

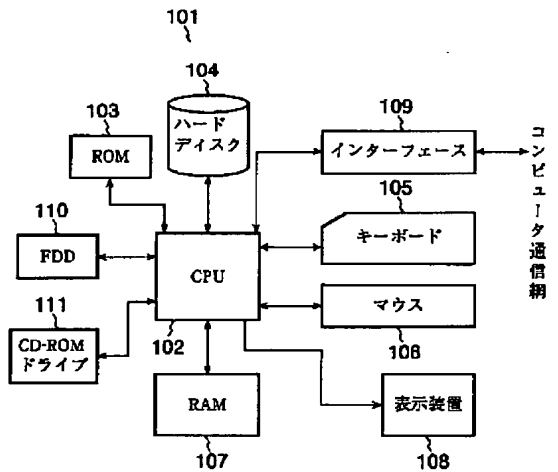
【図4】入力フォームの表示例を示す説明図である。

【図5】本発明の問合せ処理の手順を示すフローチャートである。

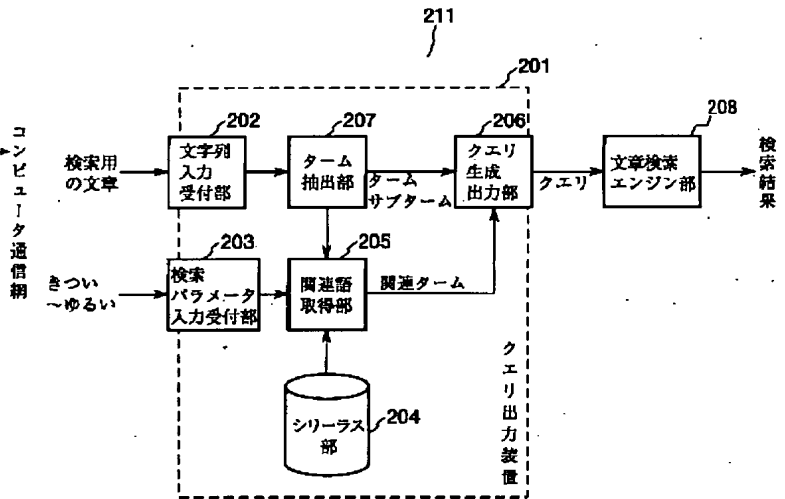
【符号の説明】

- 101 情報処理装置
- 102 CPU
- 103 ROM
- 104 ハードディスク
- 105 キーボード
- 106 マウス
- 107 RAM
- 108 表示装置
- 109 インターフェース
- 110 FDドライブ
- 111 CD-ROMドライブ
- 201 クエリ出力装置
- 202 文字列入力受付部
- 203 検索パラメータ入力受付部
- 204 シソーラス部
- 205 関連語取得部
- 206 クエリ生成出力部
- 207 ターム抽出部
- 208 文書検索エンジン部
- 211 文書検索装置
- 401 入力フォーム
- 402 文章入力欄
- 403 パラメータ用スライダ
- 404 検索実行ボタン

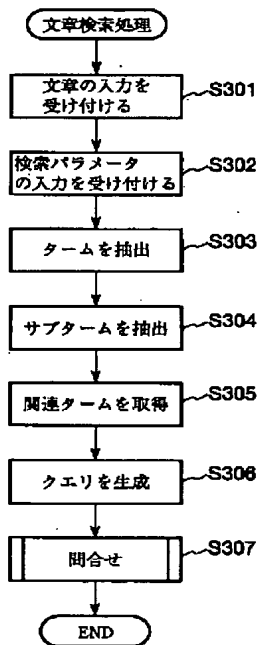
【図1】



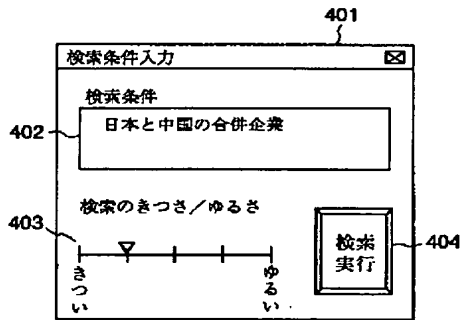
【図2】



【図3】



【図4】



【図5】

